DOCUMENT RESUME

ED 037 182 24 HE 001 425

AUTHOR	Rock, Donald A.; And Others
TITLE	The Identification and Evaluation of College Effects on Student Achievement. Final Report.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-8-0907
PUB DATE	Jan 69
GRANT	OEG-0-8-080907-4708(099)
NOTE	23p.
AVAILABLE FROM	Educational Testing Service, Princeton, New Jersey
EDRS PRICE	EDRS Price MF-$0.25 HC Not Available from EDRS.
DESCRIPTORS	*Academic Achievement, *Achievement Tests, College Environment, Comparative Analysis, Educational Resources, *Higher Education, *Learning, Learning Experience, *Predictive Measurement, Resources

ABSTRACT

This report is the result of a research effort that tried to find out what determines how much a student learns during his 4 years in college. The major purpose was to find partial answers to two basic questions. (1) If the input with respect to student ability is held constant, will identifiable groups of colleges have graduates showing greater gain in achievement than others? (2) Contingent on demonstrating differential gains between colleges, what are the characteristics of the most and least effective schools? The control variables were the verbal and mathematical scores of the SAT and the student's major field of study. The output performance variables were the area tests of the GRE Institutional Testing Program. The latter are considered achievement tests of institutional effectiveness. Institutional resources were also considered. Most of the colleges in the sample were small and included many types of liberal arts institutions. Results indicated that 85% to 91% of the between college variance was predictable from student input. A small but significant proportion was predictable from income per student, the proportion of faculty with a doctorate, full time equivalent, and the interaction of these 3 variables for all but the GRE-Social Science. (AF)

BR 80907
PA 24
OE/BR

The Identification and Evaluation of College Effects on Student Achievement

ED037182

Donald A. Rock, John A. Centra
and
Robert L. Linn

HE 001 425

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Educational Testing Service
Princeton, New Jersey

January 1969

Final Report Pursuant to Grant Number OEO-0-8-080907-4708(099) from the Office of Education, U. S. Department of Health, Education and Welfare.

The Identification and Evaluation of College Effects on Student Achievement

As greater numbers of young people continue on to college, it becomes of increasing concern to know what determines how much a student learns during his four years in college. Such information is important not only to the theorist who is attempting to understand how and to what extent college characteristics influence student behavior, but to the college administrator who requires such information for decisions concerning the optimal allocation of limited funds among many competing educational programs and processes. In addition, the recent increase in student population has been accompanied by an ever increasing flow of both public and private funds into the college system, resulting in an increasing need to evaluate the potential payoff of differential funding policies.

Many of the differences among colleges with respect to their resources have been documented by Astin and Holland (1962), Cartter (1964), and the College Data Bank of Columbia's Bureau of Applied Social Research (1966). However, little additional light has been shed on whether or not these differences produce different effects on students. Certainly any study of the impact of various colleges on students must take into account differences in students who choose to attend particular colleges. Failure to account for student talent at the time of college entrance, for example, was a criticism of the well-known studies of Knapp and Goodrich (1952) and of Knapp and Greenbaum (1953), who attempted to identify highly productive institutions by using as criteria the number of advanced graduate degrees and other scholarly rewards attained by a given institution's graduates.

Using scores on the National Merit Scholarship Qualifying Test as a control of academic ability prior to college and a sample of National Merit Scholars, Nichols (1964) and Astin (1968) found little relationship between institutional characteristics and student academic growth in college. Nichols employed a sample of 356 students at 91 colleges and used the Graduate Record Examination's (GRE) Aptitude Test as the criterion variable. On the basis of 669 students at 38 colleges, Astin more recently concluded that "traditional indices of institutional quality do not appear to contribute to student achievement" (1968, p. 661). Several factors should be considered in evaluating the conclusion reached by these two studies. First, the small sample size and the restriction to National Merit Scholars only would appear to be less than desirable for generalization. And second, because of the small number of students from each institution, both studies used individual students as the unit of analysis rather than institutional mean scores. Thus, Astin's independent effects of colleges appear quite small since he presented them as a percentage of the total individual variance after adjustment for input rather than as the percentage of the between school variance adjusted for input. This use of the ratios of school effects to the total individual variance may be misleading in that it tends to underestimate the school effect. How great the extent of underestimation is, of course, a function of the proportion of total variance which is accounted for by the between school variance. Finally, the procedure used to estimate the school effect provides relatively conservative estimates (Werts, and Linn, 1968).

This study attempted to overcome some of the handicaps characterizing the Nichols and Astin studies by (1) selecting a larger sample of colleges characterized by a wider range of ability, (2) using the institution as the sampling unit and thus, partitioning the between school variance rather than the total

individual variance, and (3) employing several different methodological approaches.

The major purpose of this research was to attempt to find partial answers to two basic questions:

(1) If the input with respect to student ability is held constant, will identifiable groups of colleges have graduates showing greater gain in achievement than others, and

(2) Contingent on demonstrating differential gains between colleges, what are the characteristics of the most and least effective schools?

Method

The input or control variables were the Verbal and Mathematical scores of the Scholastic Aptitude Tests (SAT) and the student's major field of study. The SAT was required or recommended for admission by each institution in the sample. The output performance variables were the Area Tests of the Graduate Record Examination (GRE) Institutional Testing Program. Each of the tests, i.e., Social Science, Humanities, and Natural Science, is 75 minutes in length and is intended to assess the student's grasp of basic concepts plus his ability to apply them to the variety of types of material which are presented for his interpretation (Lannholm, 1955). Thus the Area Tests are considered achievement tests of institutional effectiveness in these principal areas of learning. As an institutional measure, the tests are generally given to seniors; colleges that did not give the examination to all available seniors (or at least to all members of a designated group, such as liberal arts majors) were not included in this study.

The college descriptive measures, taken from several sources, included:

(1) measures of "institutional resources," specifically a decile ranking

of the number of books, books per student, income per student, faculty per student, and proportion of faculty with a doctorate; also full time (equivalent) undergraduate enrollment, per student expenditures, type of control, percentage of students graduating in four years, and the percentage of graduates continuing to graduate or professional schools;[1]

(2) estimated freshman orientation measures (Astin, 1965), including intellectualism, estheticism, status, pragmatism, and masculinity;

(3) college orientation measures according to Astin (1965), including realistic, scientific, social, conventional, enterprising, and artistic;

(4) average faculty compensation, and compensation per student as reported in the AAUP Journal (1968).

Only the group of characteristics under (1) was used in the majority of the analyses because groups (2), (3), and (4) were unavailable for a number of colleges.

The sample included 95 colleges that administered the GRE Area Tests in 1967 or 1968. Most college descriptive measures in group (1) above were available for 93 of these colleges. The 95 colleges also required or recommended applicants to submit the SAT for entrance. From each of these colleges, a random sample of approximately 100 seniors who had completed the GRE Area Tests was selected. For colleges with fewer than 100 seniors, the entire class was chosen. The ETS test files were then searched for the SAT scores

---

[1]The last two variables were taken from Cass and Birnbaum (1968). The other "institutional resources" variables were compiled by Columbia's Bureau of Applied Social Research (1966) and based on 1963-64 ACE and USOE Institutional data.

for these students, resulting in a final sample of 6855. This represented 74% of the 9245 students selected in the GRE sampling. The majority of SAT scores were found in either the 1963 or 1964 file years, although some were found in 1962 and 1965. Searches were not conducted beyond those years.

The institutions in this study were largely private, only four being state colleges or universities. In general, student enrollment figures were modest; only ten had more than 2000 undergraduates, none or which approached the large multiversity enrollments typified by some state and city universities. In addition to the public sector, the elitest private colleges of the Northeast were also under-represented. Approximately half of the sample was at least loosely denominational, with this group divided about equally between Catholic and Protestant denominations. In sum, the sample, while not representative of all American higher education, at least included the many types of small liberal arts institutions.

A computer based procedure developed by Rock, Barone and Linn (1968) was then used to form taxonomic groupings of colleges according to their relative profile similarity with respect to the descriptive characteristics. This system used an iterative procedure in an attempt to maximize two objective functions, one of which (the predictive objective function) is associated with the input - output matrix and the second, called the grouping objective function, yields an indication of the similarity of profiles among colleges within any one group or groups formed on the college descriptive variables or some subset of these descriptive variables. The predictive objective function in this case attempted to maximize the between group variance of the residuals (i.e., the mean predicted output subtracted from the mean observed output within each of the homogeneous groups of colleges). That is, the computer procedure provided a means for searching for that subset

of descriptive characteristics from the total set which yields groups which maximize the above predictive objective function. The direction and size of these mean residuals indicated the relative gain or loss in achievement for any one cluster of colleges when the input was held constant.

In addition to the above analyses, colleges with large positive or large negative deviations from the regression surface were compared for systematic differences on such characteristics as type of control, location and religious affiliation.

Results

In Table 1 the means, standard deviations, and intercorrelations among SAT, GRE Area Tests, and major area are reported for the sample of 6855 seniors. For each of the GRE Area Tests either SAT-V or SAT-M correlated at least .64 or higher. These correlations are somewhat higher than the correlations between the National Merit Scholarship Qualifying Tests and the GRE Area Tests reported by Astin (1968). It should also be noted that there was a positive correlation between major field of study and the appropriate GRE Area Test, suggesting that major field should be taken into account when the output scores are adjusted for input.

All of the subsequent analyses used the college as the sampling unit and thus it is the between college variance that was analyzed rather than the total variance. It would seem that the analysis of between college variance is more relevant than the analysis of total individual variance since the primary concern is the identification of college characteristics which distinguish between colleges with high and low output with input controlled. Any analysis of the total individual variance makes the implicit assumption that the college effect can be measured within college. It is also assumed

that something is known about the extent and direction of the college effect on the heterogeneity of the within college variance. It could be argued that the effect of college would reduce the heterogeneity of the within college variance. However, in the absence of empirical evidence, just the opposite might also be argued. Given this state of uncertainty it seems preferable to use only the between college variance.

In Table 2 the intercorrelations among the college means based on students with SAT scores at the 93 colleges with descriptive data are presented. As can be seen, the correlations between input (SAT mean) and output (GRE mean) are quite high. SAT-V means correlate .89 with GRE-Humanities means, and SAT-M means correlate .91, .92, and .93 with GRE-Social Science, Natural Science and Total respectively. Substantial correlations were also obtained for percentage of students majoring in Social Sciences and GRE-Social Science means ($r = .35$), percentage of students majoring in Humanities and GRE-Humanities means ($r = .51$), and percentage of students majoring in Natural Sciences and GRE-Natural Science means ($r = .37$).

The correlations between the primary college descriptive characteristics and GRE Area Test means and SAT means are reported in Table 3. Income per student and proportion of faculty with doctorates had consistently high correlations for all three Area tests and for the GRE total. The faculty compensation variables were highly correlated with the GRE and SAT means; however these data were available for a limited number of colleges.

Since residual scores were to be used for many of the analyses, an attempt was made to estimate the stability of the residuals. The sample of students within each college was randomly divided into two subsamples and GRE and SAT means were computed for each subsample. The correlations between the means for one subsample and their counterparts in the second subsample

are reported in Table 4. These correlations ranged from a low of .95 for GRE-Social Science and SAT-M to a high of .97 for GRE-Natural Science and GRE-Total, indicating a high degree of stability for the college means.

Of greater relevance are the correlations among the GRE residuals for subsample 1 with the corresponding residuals for subsample 2 when one of the SAT scores was used as a predictor. These correlations between the residuals are reported in Table 5. The least stable residual was the GRE-Social Science adjusted for SAT-M (r = .62) and the most stable residual was the GRE-Humanities adjusted for SAT-V (r = .90). In general, the residuals showed considerable stability, certainly sufficient to justify relating college characteristics to the residuals.

The multiple correlations of SAT means and proportion in major field with each of the GRE Area Test means are reported in Table 6. The multiple correlations ranged from .92 for Natural Sciences to .95 for GRE Total. The squared multiple correlations indicate the proportion of the between college output variance that can be predicted from SAT means and proportion in major. These squared multiple correlations ranged from .85 to .91 and thus approximately 9 to 15 percent of the between college output variance could not be predicted from the input measures.

Using the computer based moderated regression procedure which was described above, a subset of the college characteristics was selected which maximized the objective function having to do with the between group variance of residuals. Table 7 presents the means of the selected college characteristics and associated mean residuals for each group of colleges on each of the GRE measures. Group 1 included 54 colleges characterized by relatively high income per student and a large proportion of faculty with doctorates. This group had positive mean residuals on all three Area tests and the total. Group 2, which was comprised

of 29 colleges, had relatively low income per student and a relatively small proportion of faculty with doctorates. The mean residuals for group 2 were the largest negative residuals in all three areas and the total. Group 3, with an N of 10 colleges, was characterized by relatively low income per student and large proportion of faculty with doctorates. Group 3 had the largest positive mean residual for Social Science but negative residuals for Humanities, Natural Science and Total.

Inspection of Table 7 suggests that income per student differentiated group 1 from groups 2 and 3 while proportion of faculty with doctorate differentiated groups 1 and 3 from group 2. This combination of income per student and proportion of faculty with doctorate corresponds to an apparent interaction that was observed for GRE Social Science. That is, for GRE Social Science, colleges with low income per student can be distinguished by what the income was spent on. In short, it appears that low income colleges that spent money on obtaining a high proportion of faculty with doctorates did better in Social Sciences than those that spent their money elsewhere. In Humanities, Natural Science and Total, however, income per student appears to be the overriding consideration.

Per student expenditures were also investigated but unlike income per student, it did not discriminate between the more effective and less effective schools. The per student expenditure information was obtained from colleges on an Office of Education form 2000 and consisted of a weighted composite of the following items: 1) general administration and general expense, 2) instruction and departmental research, 3) libraries, and 4) the operation and maintenance of the physical plant. Assuming these to be accurately and uniformly reported by each college, one possible reason for its ineffectiveness is that only one of the four specific expenditures (instruction

.41 for GRE-Natural Science. With the exception of GRE-Social Science, the multiple partial correlations are statistically significant ($p < .05$). The variables with the largest weights for all four criteria were F.T.E. and I/S x F.T.E.

Astin's (1965) freshman orientation and college orientation measures were also investigated. While some of these measures (particularily Selectivity) have high zero order correlation with GRE Area test means, they were not found to be very useful in predicting the residual output measures largely because of the high correlations with the input measures. These variables as with the faculty compensation variables were investigated to only a limited extent in the present study due to the fact that they were unavailable for a number of colleges in the study. Other variables which were considered but did not aid in the prediction of the residuals were location, type of control, religious affiliation, and co-educational versus male or female institutions.

Viewing the results of the present study, several limitations should be considered. Since the sample was limited to colleges requiring both the GRE-Area Tests and the SAT, it cannot be construed as being representative of the total population of colleges. In particular, certain variables such as size, type of control, and geographic location were restricted by the availability of data. As noted earlier there were relatively few large universities, state supported institutions, or engineering colleges.

An even more serious restriction is the narrow nature of the criterion used as a measure of quality. Certainly there are many other outputs which should be evaluated in addition to achievement as measured by the GRE-Area Tests. But though the Area Tests measure only a narrow aspect of quality, the fact that these colleges choose to use the GRE-Area Tests suggests that they are relevant to the general educational goals of these institutions.

In addition to effects that a college may have on mean student achievement, colleges might be differentially effective with different types of students. For example, two colleges might have equal mean residuals yet one college might achieve this with small gains for below average students and large gains for above average students, whereas the other college might achieve this with just the opposite pattern. Such within college effects are beyond the scope of this study but are being pursued in further research.

Another limitation of this study is due to the limited nature of the college measures that were investigated. More refined measures of income, expenditures, and faculty characteristics would seem to be desirable. Variables more directly concerned with the extent and nature of student-faculty interactions would also seem to be particularly relevant.

Conclusions:

In this study, 85 to 91 percent of the between college variance was predictable from student input. A small but significant proportion of the 9 to 15 percent remaining between college variance was predictable from income per student, the proportion of faculty with a doctorate, full time equivalent and the interaction of these three variables for all but GRE-Social Science. The extent of these effects was larger for the GRE-Natural Science, Humanities, and Total than for GRE-Social Science.

While the present study analyzed the between-college variance rather than the total individual variance and used methodology (multiple partial correlation) which is more sensitive to the possibility of isolating college effects when there is a high correlation between such effects and inputs, the results were not overly encouraging. Although the college effects appear somewhat larger than in previous studies of Nichols (1964) and Astin (1968), the increments are of limited practical significance.

References

American Association of University Professors, The Economic Status of the Profession, 1966-1967. Summer 1968, 54, 2.

Astin, A. W. Undergraduate achievement and institutional "excellence." Science, 1968, 161, 661-668.

Astin, A. W. Who Goes Where to College? Chicago: Science Research Associates, 1965.

Astin, A. W., and Holland, J. L. The distribution of wealth in higher education. College and University, 1962, 37, 113-125.

Cartter, A. M. ed. American Universities and Colleges, 9th edition American Council on Education, Washington, D. C., 1964.

Cass, J., and Birnbaum, M. Comparative Guide to American Colleges and Universities, Harper & Row, 1968.

College Data Bank Code Book, Bureau of Applied Social Research, Columbia University, 1966.

Knapp, R. H., and Goodrich, H. B. Origins of American scientists. Chicago: University of Chicago Press, 1952.

Knapp, R. H., and Greenbaum, J. J. The younger American scholar, his collegiate origins. Chicago: The University of Chicago Press, 1953.

Lannholm, G. V. Development of a college level general achievement battery. Paper presented at the meetings of the American Educational Research Association, St. Louis, March 1, 1955.

Nichols, R. C. Effects of various college characteristics on student aptitude test scores. Journal of Educational Psychology, 1964, 55, 45-54.

Rock, D. A., Barone, J., and Linn, R. L. A moderated stepwise prediction system. Educational and Psychological Measurement, 1967, 3, 709-713.

Werts, C. E., and Linn, R. L. Analyzing school effects: how to use the same data to support different hypotheses. Research Bulletin 68-60. Princeton, N. J., Educational Testing Service, 1968.

Table 1

Correlations of Individual SAT Scores,
Major Area and Individual GRE Area Test Scores
N=6855

	GRE				SAT		Major		
	Soc. Sci.	Hum.	Nat. Sci.	Tot.	Verb	Math	Soc. Sci.	Hum.	Nat. Sci.
GRE-SS	1.0								
GRE-H	.53	1.0							
GRE-NS	.60	.46	1.0						
GRE-T	.86	.79	.84	1.0					
SAT-V	.65	.68	.59	.77	1.0				
SAT-M	.51	.38	.64	.62	.59	1.0			
Major SS	.26	-.05	-.06	.06	.00	-.03	1.0		
Major H	-.06	.34	-.10	.07	.17	-.04	-.40	1.0	
Major NS	-.00	-.02	.39	.15	.07	.27	-.33	-.28	1.0
Mean	484	507	506	1497	500	513	.32	.25	.19
Standard Deviation	100	97	100	249	100	105	.47	.43	.39

Table 2

Correlations of SAT-Means,

Percentage in Major Areas and GRE Area Test Means

(N = 95 colleges)

	GRE				SAT		Major		
	Soc. Sci.	Hum.	Nat. Sci.	Tot.	Verb	Math	Soc. Sci.	Hum.	Nat. Sci
GRE-SS	1.0								
GRE-H	.69	1.0							
GRE-NS	.88	.66	1.0						
GRE-T	.94	.85	.93	1.0					
SAT-V	.83	.89	.78	.91	1.0				
SAT-M	.91	.71	.92	.93	.89	1.0			
Major SS	.35	.08	.20	.23	.13	.20	1.0		
Major H	.02	.51	-.05	.17	.31	.04	-.19	1.0	
Major NS	.20	.03	.37	.22	.17	.29	-.09	-.34	1.0

Table 3

Correlations between College Descriptive Characteristics and GRE Area Test Means and SAT Means

College Characteristic	Number of College	GRE Soc. Sci.	GRE Hum.	GRE Nat. Sci.	GRE Tot.	SAT V	SAT M
umber of Books	93	.42	.15	.41	.36	.29	.41
ooks/Student	93	.09	.17	.13	.14	.15	.10
ncome/Student	93	.32	.33	.38	.38	.34	.35
aculty/Student	93	.10	.26	.03	.14	.22	.07
rop. Fac. with Doctorates	93	.37	.35	.45	.43	.34	.42
ull Time Equivalent	93	.16	-.06	.08	.07	.07	.14
er Student Expenditures	93	.41	.28	.30	.36	.38	.41
Grad. in 4 Yrs.	68	.19	.19	.12	.18	.27	.28
Going to Grad. School	82	.48	.06	.46	.37	.22	.43
verage Faculty Compensation	47	.60	.19	.47	.50	.55	.63
aculty Compensation per Student	38	.39	.41	.41	.46	.60	.58

Table 4

Correlations of Means for Subsample 1

with Means for Subsample 2

Variable	Correlations (N=95)
SS	.95
H	.96
NS	.97
Tot	.97
SAT-V	.96
SAT-M	.95

Table 5

Correlations of Residuals for Subsample 1

with Residuals for Subsample 2

Variable	Predictor	Correlation (N=95)
SS	M	.62
H	V	.90
NS	M	.75
Tot	V	.88
Tot	M	.74

Table 6

Multiple Correlations of Mean SAT-V, SAT-M and Proportion in Major Field with GRE Area Test Means (N=95 Colleges)

GRE Test	Predictors	Standard Regression Weight	Multiple Correlation
Soc. Sci.	SAT-V	.165	.93
	SAT-M	.731	
	Prop. Majoring in S.S.	.180	
Humanities	SAT-V	.968	.92
	SAT-M	-.189	
	Prop. Majoring in Hum.	.254	
Nat. Sci.	SAT-V	-.177	.94
	SAT-M	1.068	
	Prop. Majoring in N.S.	.073	
Total	SAT-V	.427	.95
	SAT-M	.553	

Table 7

Group Means on Selected College Characteristics

and GRE Residuals

Group	N	Income/ Student	Prop. Fac. with Doc.	Mean Residuals: Soc. Sci.	Hum.	Nat. Sci.	Total
1	54	7.63	6.56	.64	2.33	3.29	7.14
2	29	3.66	3.21	-2.00	-3.78	-5.02	-10.17
3	10	2.60	7.70	2.31	-1.61	-3.20	-9.04

Table 8

Multiple Partial Correlations and Standard Regression Weights for Predicting Residual Output Means from College Characteristics with Input Partialled Out

Residual Output Measure	(1) I/S	(2) F.D.	(3) F.T.E.	(4) (1 x 2)	(5) (1 x 3)	(6) (2 x 3)	Multiple Correlation
Soc. Sci.	.44	-.09	.55	-.26	-.72	.39	.28
Hum.	.21	.20	.59	.10	-.72	.00	.39
Nat. Sci.	.53	.00	.80	-.02	-1.17	.35	.41
Total	.42	.00	.73	.06	-1.00	.23	.37